

# KSHITIZ TIWARI

+1(479) 332-8248 ◊ Fayetteville, AR

[ktiware@uark.edu](mailto:ktiware@uark.edu) ◊ [linkedin.com/in/zthsk](https://www.linkedin.com/in/zthsk) ◊ [www.zthsk.github.io](http://www.zthsk.github.io)

---

## SUMMARY

Ph.D. candidate in Computer Science specializing in controllable large language models, causal modeling in NLP, and robust text generation. Published researcher (Entropy 2024, ACL-IJCNLP 2022) with expertise in adversarial robustness, parameter-efficient fine-tuning (LoRA/QLoRA), and multi-objective optimization for counterfactual generation and alignment.

---

## EDUCATION

**Ph.D. in Computer Science**, University of Arkansas Expected Dec 2026  
Research Focus: Large Language Models, Causal Inference in NLP, Robust Text Generation

**M.S. in Computer Science**, University of Arkansas Dec 2024

---

## EXPERIENCE

**Graduate Research Assistant** Aug 2021 – Present  
University of Arkansas *Fayetteville, AR*

- Conducted research on robustness, causal modeling, and controllable generation in NLP, progressing from spurious correlation mitigation in classifiers to counterfactual generation in large language models.
- **Robust Hate Speech Detection (ACL-IJCNLP 2022)**: Developed a BERT-based hate speech classifier with entropy-based regularization to mitigate spurious correlations. Improved robustness under adversarial token substitutions and reduced false positives across domains.
- **Adversarial Analysis using Minimum Description Length (Entropy, 2024)**: Proposed an information-theoretic framework to detect adversarial perturbations by estimating changes in algorithmic information via masked language modeling. Demonstrated how minimal edits alter model uncertainty and improved adversarial detection performance.
- **Counterfactual Text Generation with Large Language Models (Ongoing PhD Research)**: Designed parameter-efficient fine-tuning strategies (LoRA, soft prompting, QLoRA) to generate controlled counterfactual text (Modern ↔ Shakespearean). Implemented multi-objective optimization combining contrastive learning, SBERT-based semantic regularization, and auxiliary style classifier supervision to balance stylistic transformation with semantic preservation.
- Developed a structured evaluation framework integrating embedding-based similarity (SBERT cosine, BERTScore), classifier-based style accuracy, validity filtering, Oracle@K selection, and LLM-as-a-judge rubric scoring. Performed controlled experiments analyzing trade-offs between stylistic alignment and semantic fidelity.

---

## PUBLICATIONS

- Tiwari, K.; Zhang, L. *Implications of Minimum Description Length for Adversarial Attack in NLP*. Entropy, 2024.
- Tiwari, K.; Yuan, S.; Zhang, L. *Robust Hate Speech Detection via Mitigating Spurious Correlations*. ACL-IJCNLP, 2022.

---

## SELECTED PROJECTS

**LLM-Based Resume Screening and Candidate Ranking**: Designed an end-to-end pipeline to parse PDF resumes into structured representations, generate semantic summaries using LLM prompting, and rank candidates via SBERT similarity scoring against job descriptions. Implemented composite ranking metrics for top-N retrieval.

**Multimodal Hate Meme Detection:** Developed a multimodal classification system combining Vision Transformers and BERT representations for joint image-text reasoning. Integrated LLM prompting for enhanced cross-modal alignment and improved benchmark performance.

**Semantic Search Benchmarking:** Benchmarked classical topic modeling (LSA, LDA) against transformer-based embeddings (SBERT), demonstrating improved retrieval performance over GloVe baselines.

## TECHNICAL SKILLS

---

**LLM Fine-Tuning & Alignment:** LoRA, QLoRA, PEFT, Instruction Tuning, Contrastive Learning, Auxiliary Classifier Supervision, LLM-as-a-Judge Evaluation

**NLP & Deep Learning:** PyTorch, Hugging Face Transformers, BERT, SBERT, Vision Transformers, Scikit-learn

**Evaluation & Experimentation:** BERTScore, Embedding Similarity, Ranking Metrics (Oracle@K), JSON Pipelines, LaTeX Reporting

**Programming:** Python (primary), C++, SQL